

The Contribution of Interchain Salt Bridges to Triple-Helical Stability in Collagen

Thomas Gurry,[†] Paul S. Nerenberg,[‡] and Collin M. Stultz^{§*}

[†]Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, United Kingdom; [‡]Department of Physics and Research Laboratory for Electronics; and [§]Division of Health Sciences and Technology, Department of Electrical Engineering and Computer Science and Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT Studies on collagen and collagen-like peptides suggest that triple-helical stability can vary along the amino acid chain. In this regard, it has been shown that lysine residues in the Y position and acidic residues in the X' position of (GPO)₃ GXYGX'Y'(GPO)₃ peptides lead to triple-helical structures with melting temperatures similar to (GPO)₈ (where O is hydroxyproline), which is generally regarded as the most stable collagen-like sequence of this length. This enhanced stability has been attributed to the formation of salt bridges between adjacent collagen chains. In this study, we explore the relationship between interchain salt bridge formation and triple-helical stability using detailed molecular simulations. Although our results confirm that salt bridges promote triple-helical stability, we find that not all salt bridges are created equal. In particular, lysine-glutamate salt bridges are most stabilizing when formed between residues in the middle strand (B) and the trailing strand (C), whereas lysine-aspartate salt bridges are most stabilizing when formed between residues in the leading (A) and middle (B) strand—the latter observation being consistent with recent NMR data on a heterotrimeric model peptide. Overall, we believe these data clarify the role of salt bridges in modulating triple-helical stability and can be used to guide the design of collagen-like peptides that have specific interchain interactions.

INTRODUCTION

The collagens form a family of extracellular matrix proteins that play important roles in maintaining the structural integrity of a number of tissues including blood vessels, bone, ligaments, and tendons (1,2). Given the prevalence of collagen in a relatively large number of tissues, it is not surprising that collagen itself has been implicated in several common human diseases (1–4). Consequently, studies aimed at understanding the structure and metabolism of collagen are of particular importance.

The structure of collagen is characterized by three distinct amino acid chains that fold together to form a triple helix. This helical structure contains interchain backbone hydrogen bonds between the glycine amide protons and the carbonyl oxygens of residues in the X position of the corresponding GXY triplet on the adjacent chain (5). A high abundance of prolines in the X position and hydroxyprolines in the Y position has been noted in sequenced fibrillar collagens, accounting for ~30% of residues in these positions (6). Proline residues induce a polyproline II-like helical conformation in individual strands, greatly reducing their entropy in the unfolded state and thereby the resulting loss of entropy associated with triple helix formation (7,8). The mechanism behind hydroxyproline-induced triple-helical stabilization may be more complex. It was thought initially that the hydroxyl groups helped coordinate both intra- and intermolecular hydration networks, but other models suggest that the stereoelec-

tronic effect of the hydroxyl group stabilizes a pucker of the pyrrolidine ring and backbone dihedral angles that are conducive to the formation of triple-helical structure (7,9).

Biochemical experiments in which single residue mutants were created in homotrimeric model collagen peptides containing the sequence (GPO)₃GXY(GPO)₄, where either the X or the Y residue was mutated from proline or hydroxyproline, reveal insights into the contribution of different residues to triple-helical stability (6). Compared to the highly stable control sequence of (GPO)₃GPO(GPO)₄, which had a melting temperature of 47.3°C, a lysine in position Y lowered the melting point to a T_m of 41.5°C, as did glutamate and aspartate residues in position X, which had T_m values of 42.9°C and 40.1°C, respectively (6). However, when these mutations were combined in a model peptide with the sequence (GPO)₃GXYGX'Y'(GPO)₃, where the X' and Y residues were mutated simultaneously, a striking stabilization of the triple helix was observed, with T_m values of 47.8°C and 47.1°C in the (GPO)₃GPKGEO(GPO)₃ and (GPO)₃GPKGDO(GPO)₃ mutants, respectively; i.e., stabilities comparable to the (GPO)₈ peptide (10). Clearly, interactions that are not accounted for by the additive effect of individual residue stabilization were promoting triple-helical stability in these peptides.

Additional studies suggest that interchain electrostatic interactions between lysine residues in the Y position and acidic residues in the X' position underlie this unexpected increase in stability (10). In particular, experiments carried out at a pH below the pKa of the glutamate side chain significantly reduces the melting temperature of the (GPO)₃GPKGEO(GPO)₃ peptide (10). Moreover, energy minimization of triple-helical structures in vacuo, and with limited

Submitted November 1, 2009, and accepted for publication January 19, 2010.

*Correspondence: cmstultz@mit.edu

Editor: George I. Makhatadze.

© 2010 by the Biophysical Society
0006-3495/10/06/2634/10 \$2.00

doi: 10.1016/j.bpj.2010.01.065

water molecules, suggest that intrachain salt bridges are energetically favorable as long as the oppositely charged residues are separated by at most two residues (11). Furthermore, χ^2 tests on the frequency of KGE and KGD residue pairs in fibrillar collagen sequences suggest that these residues occur more frequently together than would be expected by their individual frequencies in collagen alone, suggesting that these regions may play additional roles in modulating collagen stability (10). Despite these results, the precise energetic contribution of interchain salt bridges to triple-helical stability has not been fully explored.

To clarify the energetic contribution that interchain salt bridges have on triple-helical stability, we calculated the relative contribution of salt bridge interactions to the folding free energy of the triple-helical structure. As each amino acid chain in collagen has a precise chain stagger, three distinct salt bridge configurations are possible—all of which are examined in this work. In this manner, we present what we believe to be a comprehensive assessment of the relationship between salt bridge formation and triple-helical stability.

METHODS

Construction of the initial models

Peptides with capped ends (acetylated on the N-terminal residues and amidated on the C-terminal residues) were built using the triple-helical collagen building script (The BuScr) 1.06 and CHARMM 35b2 with the CHARMM22/CMAP all-atom parameters (12–14). The peptides were then solvated with 4946 (for the $(\text{GPO})_3\text{GPKGEO}(\text{GPO})_3$ peptide) or 4952 (for the $(\text{GPO})_3\text{GPKGDO}(\text{GPO})_3$ peptide) TIP3P water molecules using the MMTSB tool set (15). Water molecules were subjected to a cylindrical stochastic boundary potential of radius 22.5 Å and length 100 Å using the miscellaneous mean field potential with one cylindrical and two planar constraints. Before solvation, the model triple helices were brought to salt bridge configurations by restraining the distance between the lysine N ζ and the glutamate C δ or aspartate C γ , initially with a force constant of 20 kcal mol⁻¹ Å⁻² for 500 steps of steepest descent and 1000 steps of adopted basis Newton Raphson minimization, followed by 1000 more of adopted basis Newton Raphson with a force constant of 2 kcal mol⁻¹ Å⁻².

Molecular dynamics simulations

Each system was equilibrated for 250 ps followed by 10 ns of production molecular dynamics simulations using CHARMM 35b2 (13). The system was linearly heated to a temperature of 298.15 K over 50 ps during equilibration and then coupled to a Nosé-Hoover thermostat at the same temperature (16). Coordinates were saved every 2 ps. Initial simulations suggested that protein would adopt slightly bent conformations where the ends of the molecule would wander outside of the simulation cylinder. Similar observations have been made in other simulations of triple-helical peptides (17). To ensure that physiologically relevant (i.e., triple-helical) states were sampled and that the peptides remained within the solvent region for the duration of the simulations, all simulations used a harmonic constraints on the first and last three residue backbone atoms with a force constant of 2 kcal mol⁻¹ Å⁻².

Umbrella sampling

We calculated the free energy of interchain salt bridge formation using umbrella sampling (18,19). The approach outlined here is similar to what

was described in our previous work (20). The reaction coordinate was defined as the interatomic distances between the lysine N ζ and the glutamate C δ (in the case of $(\text{GPO})_3\text{GPKGEO}(\text{GPO})_3$), or between the lysine N ζ and the aspartate C γ (in the case of $(\text{GPO})_3\text{GPKGDO}(\text{GPO})_3$). The carboxyl carbons were chosen as opposed to the carboxyl oxygens because measuring interatomic distances to the carbon atoms circumvents ambiguities that arise from rotations in the C γ -C δ bond in the glutamate or the C β -C γ bond in aspartate (i.e., each carbonyl oxygen can accept a hydrogen bond from the lysine side chain). Simulations of the triple-helical peptide suggested that interatomic distances of 3.6 Å and 3.4 Å correspond to salt bridges in the $(\text{GPO})_3\text{GPKGEO}(\text{GPO})_3$ and $(\text{GPO})_3\text{GPKGDO}(\text{GPO})_3$ peptides, respectively.

Values of the reaction coordinate, ξ , were sampled starting from 2.4 Å and up to 13.0 Å in 0.2 Å increments. In practice, the starting value for these runs was 3.4 Å and sampling was done from 3.4 → 13.0 Å and 3.4 Å → 2.4 Å. A harmonic biasing potential with a force constant of 80 kcal mol⁻¹ Å⁻² was introduced at each window to bias the system toward sampling a particular value of ξ .

Each window was run for 250 ps. Trajectories were saved every 0.1 ps, resulting in 2500 frames per window. For each window we plotted the running average of the system's internal energy and fit an exponential distribution to determine the relaxation time constant, τ . Because the fits were excellent for all windows, we defined the equilibration period for a given window to be 3τ . By coupling a relatively large force constant with small increments of the reaction coordinate, the convergence time of our simulations was relatively short (i.e., $3\tau < 150$ ps for each window) and we were also able to achieve sufficient sampling overlap between neighboring windows. This allowed a production period of at least 100 ps (or 1000 frames) for each window. Each peptide retained its triple-helical conformation throughout the umbrella sampling calculations; i.e., the average backbone RMSD from the initial prototypical triple-helical state was >1 Å for all umbrella sampling windows.

The final potential of mean force, $W(\xi)$, in the i^{th} window is calculated using the expression (21):

$$W(\xi) = -\beta^{-1} \ln p^*(\xi) - V_i(\xi) + C_i, \quad (1)$$

where $p^*(\xi)$ is the probability distribution of the reaction coordinate in the biased system, $V_i(\xi)$ is the umbrella potential and C_i is a constant at window i . Although a biasing potential is used to generate the data, the pmf itself represents the free energy distribution in the unbiased system and thus the precise form of the umbrella potential does not, in principle, affect the final result (21).

Data from the production runs was combined to form the unbiased probability distribution using the WHAM algorithm (22,23). Similar to what was described above, umbrella sampling simulations used additional restraints on the backbone atoms of the first and last three terminal residues (force constant of 2 kcal mol⁻¹ Å⁻², as noted in the previous section) to ensure that we simulated states in the vicinity of the triple-helical state and that the peptides remained within the solvated stochastic boundary cylinder for the duration of the simulations. To ensure that the precise choice of the harmonic restraint did not unduly influence our results, we computed the energy associated with these terminal residue harmonic constraints for all umbrella sampling windows. Windows where the contribution from these harmonic restraints was significant (>1.0 kcal mol⁻¹) were excluded from the potential of mean force (pmf) calculation. With this convention only windows centered about values of $\xi > 12.0$ Å for the C-A interaction in $(\text{GPO})_3\text{GPKGEO}(\text{GPO})_3$ were excluded.

Throughout this work, we adopt the notation ξ_{AB} to denote the interatomic distance between the lysine on chain A and acidic residue on chain B. Similar definitions apply for ξ_{BC} and ξ_{CA} .

$\Delta\Delta G$ calculations

$\Delta\Delta G$ values were calculated as the difference in energy between the salt bridge state and the local energy minimum (or the relatively flat region of

the conformational free energy surface à la the A-B interaction in the GPKGDO peptide, described in the Results) corresponding to conformations where the salt bridge is broken. For each interaction, the enthalpy, ΔH , was estimated as the difference in internal energy of the salt bridge formed state and the salt bridge broken state. The internal energy of the salt bridge formed state was obtained by averaging over the potential energy of all snapshots taken from the umbrella sampling window corresponding to the region of the free energy surface representing the salt bridge (i.e., the global energy minimum). Similarly, the internal energy of the salt bridge broken state was obtained by averaging over the region of the free energy surface corresponding to the state where the salt bridge is broken. In all cases but one, this latter state corresponds to a local energy minimum. In the case of the A-B interaction in the GPKGDO peptide, it corresponds to the relatively flat region of the conformational free energy surface (see Results). For all interactions, these definitions correspond to the salt bridge formed and broken states observed in the 10-ns molecular dynamics trajectories. The entropic contribution, $-T\Delta S$, was defined as $\Delta\Delta G - \Delta H$.

RESULTS

In tropocollagen the different collagen chains are staggered with respect to each other resulting in asymmetric interactions between chains. Consequently, interchain salt bridges between the A-B, B-C, and C-A pairs have distinct orientations (Fig. S1 in the Supporting Material). In light of this, we examined the effect of each type of salt bridge on triple-helical stability.

Dynamical trajectories of salt bridge-containing peptides

We begin with an analysis of thermal fluctuations in interchain salt bridge-containing structures. Ten nanosecond simulations of the equilibrated (GPO)₃GPKGEO(GPO)₃ (henceforth referred to as the GPKGEO peptide) and (GPO)₃GPKGDO(GPO)₃ (henceforth referred to as the GPKGDO peptide) systems were carried out to observe the behavior of the interaction partners. During the simulation the overall triple-helical structure is maintained; i.e., after 10 ns the backbone RMSD from the initial triple-helical structure is 0.83 Å for the GPKGEO peptide and 0.80 Å for the GPKGDO peptide.

To characterize the behavior of the salt bridge pair, we use the interatomic distance between the N ζ of lysine and the C δ glutamate (in the case of the GPKGEO peptide), and the distance between the N ζ of lysine and the C γ of aspartate (in the case of GPKGDO peptide). Simulations of the GPKGEO triple-helical peptide highlight the difference between the interchain interactions (Fig. 1). The interresidue distance between the side chains of Lys and Glu in the A-B interaction mainly fluctuates between two states in GPKGEO—the first has an average distance of 3.6 Å (that corresponds to a salt bridge between the two residues), and the second has an average distance of 5.9 Å (where the salt bridge is broken). Overall, during the 10 ns simulation the B-C salt bridge is the most stable about a mean value of 3.6 Å for 74% of the 10 ns, whereas the A-B and C-A salt bridges are considerably weaker (Fig. 1).

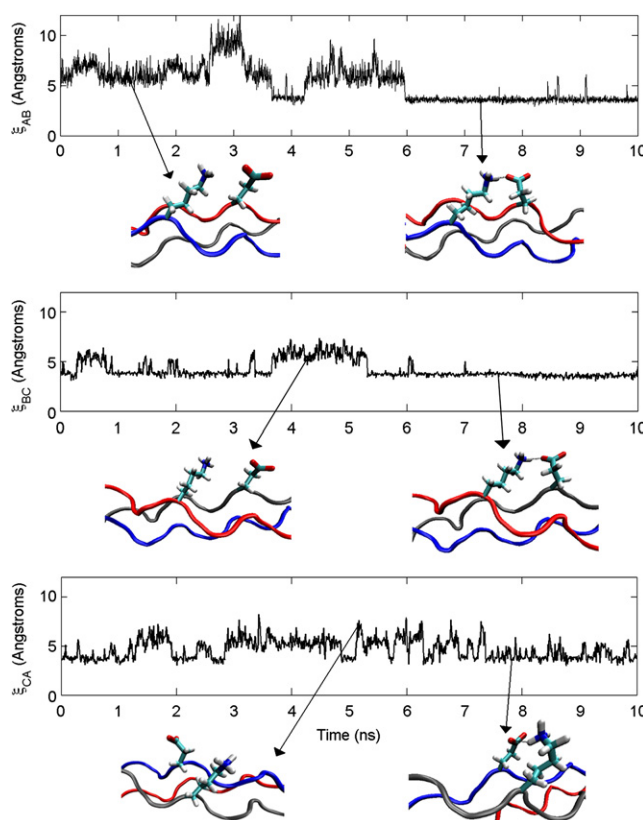


FIGURE 1 Dynamical time series data for the GPKGEO peptide. Snapshots from the various states are shown below the time series for each interchain interaction.

By contrast, in simulations of the GPKGDO peptide, the A-B salt bridge is stable during the 10 ns (formed 96% of the time) about a mean value of 3.4 Å. The B-C salt bridge is slightly less so (86% of the time), also about a mean value of 3.4 Å, and the system seems to sample a secondary state with average distance 5.8 Å, along with another state at higher values of ξ_{BC} . The C-A salt bridge is by far the least stable interaction (Fig. 2).

The effect of salt bridge formation on triple-helical stability

We wish to determine the energetic contribution of each salt bridge to triple-helical stability. In this regard, we consider two folding reactions. In the first, the protein folds from an unfolded state to a triple-helical state where the salt bridge of interest is formed. The associated chemical reaction is denoted by $U \rightarrow F_{\text{on}}$ where U represents the unfolded ensemble and F_{on} represents the folded triple-helical that contains the salt bridge. We note that this formalism makes no assumptions about the nature of the unfolded state in that the unfolded ensemble contains all possible configurations of the three dissociated chains where each chain samples conformers that span the full range of backbone and side-chain dihedrals that are consistent with the fully

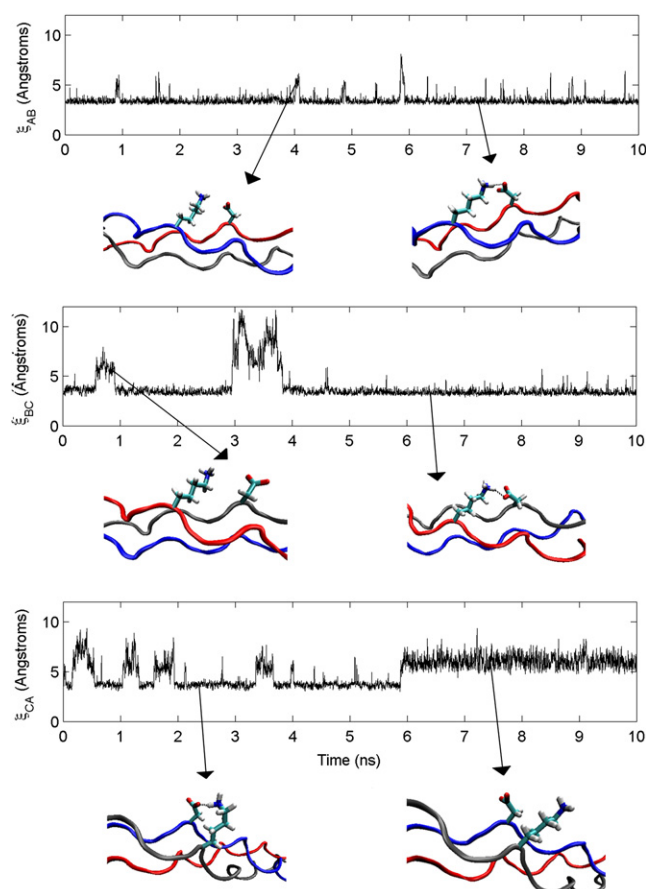


FIGURE 2 Dynamical time series data for the GPKGDO peptide. Snapshots from the various states are shown below the time series for each inter-chain interaction.

unfolded state. Similarly, the second reaction, $U \rightarrow F_{\text{off}}$, corresponds to the protein folding from the fully unfolded state to a triple-helical state where the salt bridge is broken. Because the beginning reactions of both reactions are identical, we use the thermodynamic cycle shown in Fig. 3 to quantitatively assess the effect of interchain salt bridge formation on triple-helical stability. In path 1, the three collagen chains fold to a state that contains a salt bridge ($\Delta G_{\text{fold}}^{\text{on}}$ is the associated free energy change), and in path 2 folding leads to a triple-helical structure where the salt bridge is broken ($\Delta G_{\text{fold}}^{\text{off}}$ is the associated free energy change). The contribution of the salt bridge to triple-helical stability is given by the difference $\Delta\Delta G = \Delta G_{\text{fold}}^{\text{on}} - \Delta G_{\text{fold}}^{\text{off}}$. In principle, one could carry out simulations to directly compute $\Delta G_{\text{fold}}^{\text{on}}$ and $\Delta G_{\text{fold}}^{\text{off}}$, however the direct calculation of the free energy associated with such folding reactions is computationally demanding. This is especially true for this system because simulating the unfolded state would require building a model for the fully unfolded state. However, as the thermodynamic path shown in Fig. 3 shows, the effect of salt bridge formation on protein stability is equivalent to the free energy associated with salt bridge formation in the folded state; i.e.,

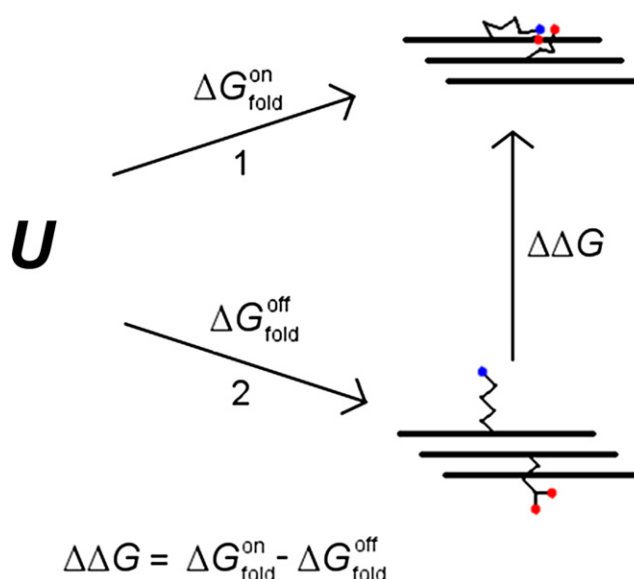


FIGURE 3 Thermodynamic cycle of collagen triple helix formation involving the electrostatic interactions under study. The unfolded state (U , left) represents the ensemble of all conformations of three dissociated chains, where each individual chain is free sample main-chain and side-chain conformations that are consistent with the fully unfolded state. The folded conformations (right) represent triple-helical conformations where the salt bridge of interest is formed (top) or broken (bottom).

$\Delta\Delta G$ can be calculated from an analysis of the folded state alone.

The free energy difference associated with salt bridge formation, $\Delta\Delta G$, could be computed using a number of methods—each of which makes a different set of assumptions. For example, one could slowly turn off the electrostatic interactions between the two side chains that form the salt bridge and use thermodynamic perturbation/integration to compute the resulting free energy change (24,25). Strictly speaking, this method constitutes one way to calculate the electrostatic contribution to $\Delta\Delta G$, however, the overall electrostatic contribution is somewhat distinct from the contribution of a salt bridge. In particular, two oppositely charged side chains that are very distant from one another would not be considered to form a salt bridge. However, because electrostatic interactions are long range in character, these well separated side chains will still make some electrostatic contribution to the overall energy of the system. Hence the broken salt bridge state does not correspond to a state where all electrostatic interactions between the side chains are turned off.

We use umbrella sampling to compute the contribution of the salt bridge to protein stability. In this approach, the beginning and end states of the reaction are defined solely by the distance between the two side chains in question and long range electrostatic interactions are considered in both the state where the salt bridge is formed and the state where it is broken. Central to this method is a clear definition

of the states corresponding to the conformation where the salt bridge is formed and the state corresponding to a broken salt bridge. To identify these end states we use the dynamical trajectories shown in Fig. 1 and Fig. 2 as a guide. In each trajectory, the system fluctuates between two distinct regions of conformational space, where the first region corresponds to a relatively low value of the reaction coordinate and that has representative structures that place the oppositely charged side chains within hydrogen bonding distance of one another. The second state corresponds to a larger value of the reaction coordinate where representative structures have side chains that are too far apart to allow direct hydrogen bonding to happen. These two states form the beginning and end states for our calculation of $\Delta\Delta G$.

The calculated free energy profiles correspond to the rigorous free energy of the system and includes both enthalpic and entropic contributions to the overall energy (21). Therefore, we begin our analysis with an investigation of the free energy profile (pmf) for each interchain salt bridge in the folded state in light of the dynamical trajectories discussed above. These insights are then used to compute quantitative estimates for $\Delta\Delta G$.

GPKGEO

The pmf for the A-B interchain interaction contains two well defined minima (Fig. 4). The global energy minimum is located at $\xi_{AB} = 3.3$ Å, and corresponds to the state where an interchain salt bridge is formed between the lysine and glutamate side chains, consistent with the state with mean value 3.6 Å in the dynamical trajectory in Fig. 1. The second minimum is located at $\xi_{AB} = 5.5$ Å, a value similar to 5.9 Å, the mean value of the secondary state in Fig. 1. The relative

free energy difference between the two states is quite modest—only 0.6 kcal/mol. This observation is in qualitative agreement with the data shown in Fig. 1. In particular, unrestrained simulations that begin with the AB salt bridge, frequently sample states where the salt bridge is broken.

The pmf for the B-C interaction also contains two energy minima, where the global energy minimum corresponds to an interchain salt bridge at $\xi_{BC} = 3.7$ Å (Fig. 5). The second minimum is more shallow and is located at $\xi_{BC} = 5.7$ Å and has a free energy that is 2.4 kcal/mol ($\sim 4 k_B T$ at room temperature) higher than the ground state. Indeed, unrestrained simulations suggest that B-C salt bridges are considerably more stable (Fig. 1).

The pmf for the C-A interaction contains two energy minima (Fig. 6). The global energy minimum is found at $\xi_{CA} = 3.7$ Å and corresponds to an interchain salt bridge. Again, the second minimum is shallow and is located at $\xi_{CA} = 4.7$ Å. The first of these two values is consistent with the mean values extracted from the trajectories in Fig. 1 (3.6 Å). The second state has a value ξ_{CA} that is somewhat lower than the average value in Fig. 1, but representative structures in both cases correspond to states where the side chains are well beyond hydrogen bond distance. The second state has free energy that is 1.3 kcal/mol higher than the ground state.

Analysis of solvent molecules within hydrogen-bonding distance of the lysine and glutamate side chains suggests that in each case, the secondary local energy minimum forms when the side chains separate enough to allow water molecules to enter the space between the side chains. For example, for the A-B interaction, at a separation of 5.5 Å a single water molecule can hydrogen bond to both side chains. An analysis

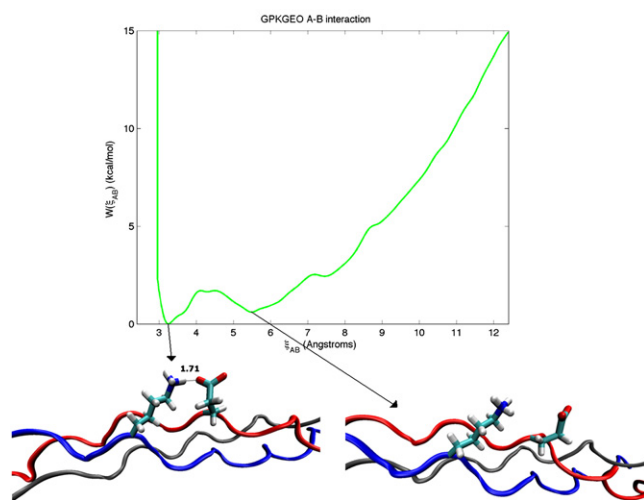


FIGURE 4 Potential of mean force constructed using the ξ_{AB} reaction coordinate for GPKGEO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{AB} = 3.4$ Å and $\xi_{AB} = 5.6$ Å, respectively, corresponding to the two energy minima of the pmf.

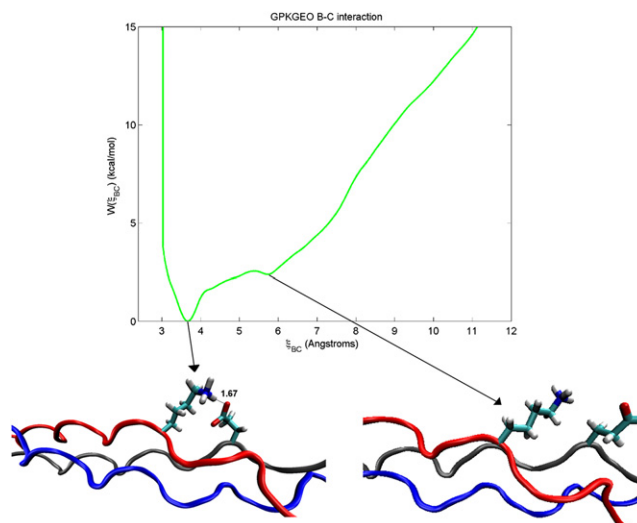


FIGURE 5 Potential of mean force constructed using the ξ_{BC} reaction coordinate for GPKGEO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{BC} = 3.8$ Å and $\xi_{BC} = 5.8$ Å, respectively, corresponding to the two energy minima of the pmf.

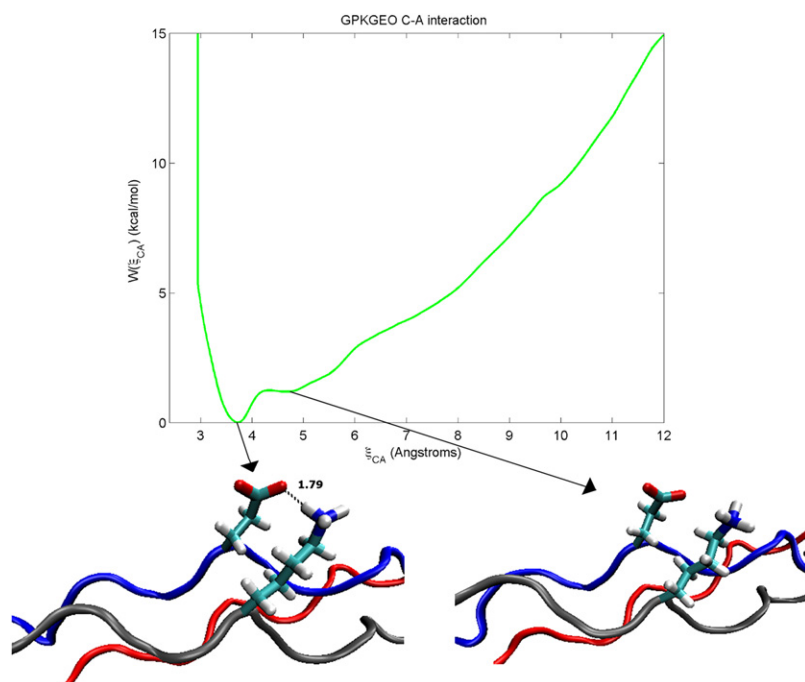


FIGURE 6 Potential of mean force constructed using the ξ_{CA} reaction coordinate for GPKGEO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{CA} = 3.8 \text{ \AA}$ and $\xi_{CA} = 4.8 \text{ \AA}$, respectively, corresponding to the two energy minima of the pmf.

of the window that is restrained to sample around $\xi_{AB} = 5.5 \text{ \AA}$, $\xi_{BC} = 5.7 \text{ \AA}$, and $\xi_{CA} = 4.7 \text{ \AA}$, different water molecules take turns hydrogen bonding to both side chains.

GPKGDO

The pmf for the A-B interaction in GPKGDO contains a prominent energy minimum located at $\xi_{AB} = 3.3 \text{ \AA}$ (Fig. 7), consistent with the value of 3.4 \AA observed in the dynamical trajectory in Fig. 2. This state corresponds to a conformation where an interchain salt bridge is formed

between the lysine and the aspartate residues. Trajectories of the triple-helical conformation infrequently sample a state having $\xi_{AB} \sim 5 \text{ \AA}$ that corresponds to a state where the salt bridge is broken (Fig. 2). This state corresponds to a shallow shelf on the conformational free energy surface that has a free energy that is 3.6 kcal/mol higher than the ground state (Fig. 7).

The pmf for the B-C interaction in GPKGDO contains two energy minima (Fig. 8). The global energy minimum is found at $\xi_{BC} = 3.3 \text{ \AA}$ (also similar to the value of 3.4 \AA in

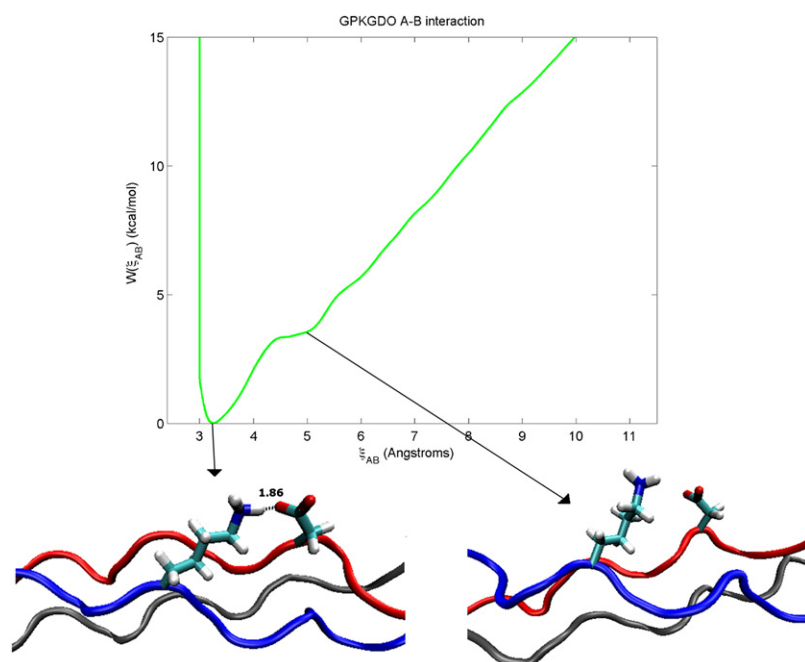


FIGURE 7 Potential of mean force constructed using the ξ_{AB} reaction coordinate for GPKGDO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{AB} = 3.4 \text{ \AA}$ and $\xi_{AB} = 5.0 \text{ \AA}$, respectively, corresponding to the two energy minima of the pmf.

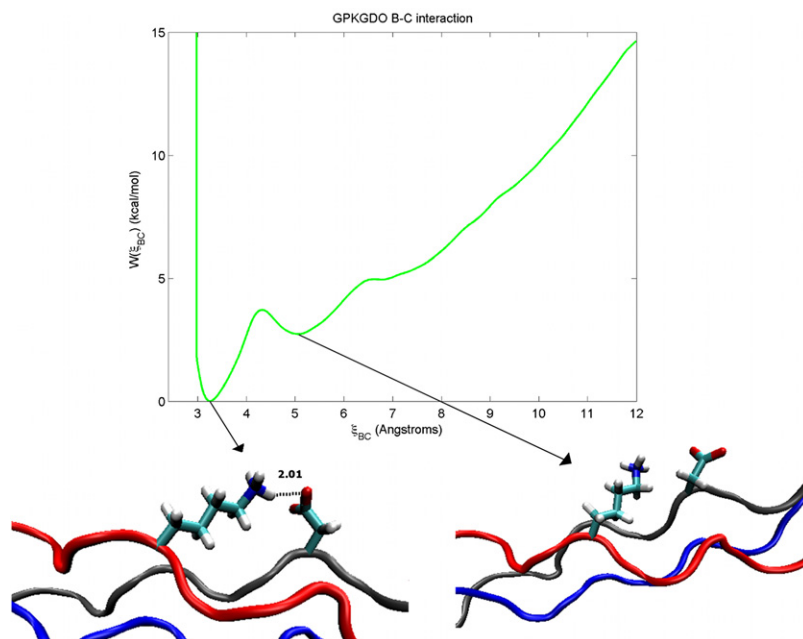


FIGURE 8 Potential of mean force constructed using the ξ_{BC} reaction coordinate for GPKGDO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{BC} = 3.4$ Å and $\xi_{BC} = 5.0$ Å, respectively, corresponding to the two energy minima of the pmf.

Fig. 2), and corresponds to an interchain salt bridge. The second energy minimum is found at $\xi_{BC} = 5.0$ Å, that has a free energy that is 2.9 kcal/mol relative to the ground state (~ 5 $k_B T$ at room temperature). Consequently, this state is infrequently sampled in dynamical trajectories that begin with the B-C salt bridge formed (Fig. 2).

Both metastable states for the A-B and B-C interaction occur at interatomic distances that allow individual water molecules to hydrogen bond to both states, in a manner similar to what was described for the GPKGEO trajectories above. For the shallow metastable state for the A-B interaction, $\sim 92\%$ of the time different water molecules again take turns hydrogen bonding to the two side chains. For the B-C metastable state, this occurs $\sim 89\%$ of the time.

The pmf for the C-A interaction in GPKGDO has two energy minima (Fig. 9). The global minimum is located at $\xi_{CA} = 3.7$ Å, which corresponds to an interchain salt bridge. In addition, a secondary minimum can be found at $\xi_{CA} = 5.4$ Å with a relative free energy of 1.1 kcal/mol. The secondary minimum is considerably broader than the secondary minimum for the other interaction pairs. Such larger distances correspond to states where several water molecules intervene, bridging the two side chains in a more complicated array. This is consistent with the dynamical time series in Fig. 2, which appears to sample states in this vicinity when not in a salt bridge (defined as having values of $\xi_{CA} > 4.2$ Å).

Calculation of $\Delta\Delta G$ values

We use the different pmfs to calculate $\Delta\Delta G$ —the contribution of each salt bridge to triple-helical stability—as outlined in Fig. 3. $\Delta\Delta G$ is then calculated as the difference in energy between the salt bridge state and the local energy minimum (or the relatively flat region of the conformational free energy

surface à la the A-B interaction in the GPKGDO peptide) corresponding to conformations where the salt bridges are broken. These values are shown in Table 1. In all cases, salt bridge formation stabilizes the triple-helical structure. However, for the GPKDEO peptide, B-C salt bridges are the most stabilizing and in the GPKGDO peptide, A-B salt bridge pairs are the most favorable.

Table 1 also decomposes the free energy of into both enthalpic, ΔH , and entropic, $-T\Delta S$, contributions. In each case, $-T\Delta S > 0$, which implies $\Delta S < 0$; i.e., there is a decrease in the entropy of the system on salt bridge formation. However, the decreased entropy is more than compensated for by the favorable enthalpic contribution.

DISCUSSION

The effect of salt bridges on protein stability has long been a subject of great interest (25–28). It has been shown that for globular proteins the contribution of salt bridges to protein stability can vary greatly depending on the system studied. For thermophilic and hyperthermophilic proteins, for example, salt bridges may on average contribute to the increased thermal stability of these systems (29). The importance of salt bridges for proteins that exist in more temperate environments is a matter of debate. There are data to suggest that salt bridge formation in these systems is often not stabilizing and when it is, it likely only marginally contributes to overall stability (25–28,30–32).

Unlike globular proteins, all side chains in the triple-helical structure of collagen are solvent exposed and collagen itself has no hydrophobic core. Therefore although studies on the contribution of salt bridges to globular proteins have advanced our understanding of protein folding in general, it

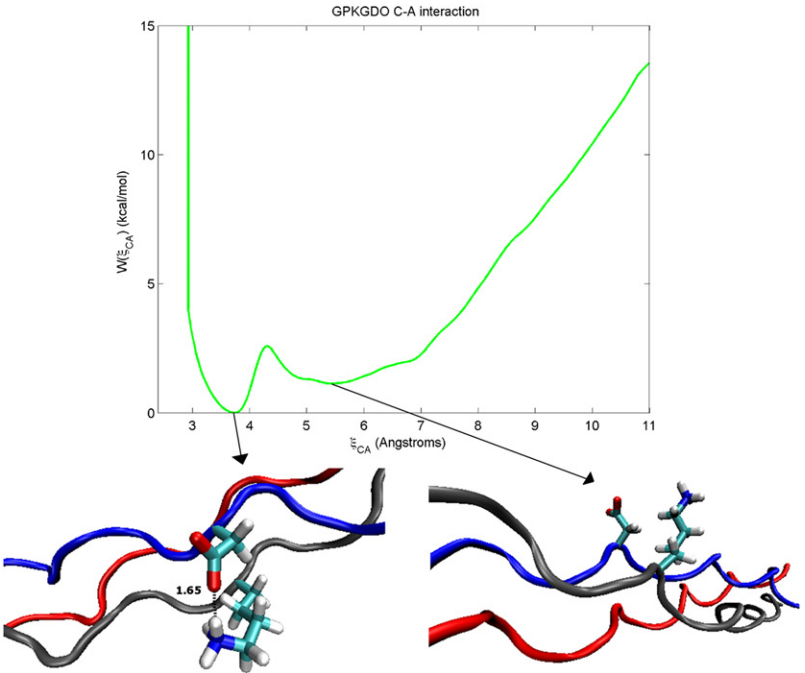


FIGURE 9 Potential of mean force constructed using the ξ_{CA} reaction coordinate for GPKGDO. Also shown are snapshots inside umbrella sampling simulation windows centered on $\xi_{CA} = 3.8 \text{ \AA}$ and $\xi_{CA} = 5.4 \text{ \AA}$, respectively, corresponding to the two energy minima of the pmf.

is not clear how these insights relate to triple-helical stability in collagen. Moreover, the effect of salt bridge formation on fibrillar proteins, like collagen, has yet to be fully explored. Previous work has shown that the introduction of single acidic or basic residues into $(\text{GPO})_3\text{GXY}(\text{GPO})_4$ host-guest peptides leads to a significant decrease in the peptide melting temperature (6). However, double mutants $(\text{GPO})_3\text{GXYGX}'\text{Y}'(\text{GPO})_3$ that contain pairs of oppositely charged residues led to an unexpected increase in protein stability, suggesting that interchain salt bridges may be an important regulator of triple-helical stability (10).

In this study we explored the energetic contribution of salt bridges to the thermodynamics of triple helix folding. First, we show that an analysis of the relative effect of salt bridge formation on triple-helical folding can be determined from an analysis of the folded state alone. Calculated conformational free energy landscapes for the folded triple-helical structure were computed using a reaction coordinate that varied the distance between the side chains of the oppositely charged residues, thereby sampling states where the salt bridge is formed and states where it is broken. The final free energy profiles represent the rigorous free energy of

the system and includes both enthalpic and entropic contributions to the overall energy (21). All of the resulting free energy profiles contain a prominent global energy minimum and a second metastable state that has a larger interresidue distance between the salt bridge pair partners. In each case, the global energy minimum and the metastable state correspond to conformers sampled in unrestrained dynamical trajectories of the salt bridge pairs. For all interaction pairs, the global energy minimum corresponds to the conformation that contains a salt bridge whereas the metastable state corresponds to an interresidue distance that enables a water molecule to sit between the oppositely charged side chains, hydrogen bonding to both.

Interestingly, we found that salt bridges can exhibit a range of energetic contributions to triple-helical stability. Some salt bridges have only a marginal effect on stability (A-B pair in GPKGEO), whereas others significantly stabilize the folded state (B-C in GPKGEO and A-B in GPKGDO). Despite the structural differences between collagen and globular proteins, we find that salt bridges can stabilize collagen by 0.6–3.6 kcal/mol—a value similar to the folding stability (1–5 kcal/mol) contributed by both buried and surface salt bridges in globular proteins (26,30–32).

The asymmetry between interchain interactions is a critical property of interchain electrostatic interactions in collagen. Both an unrestrained simulation (Fig. 2) and the $\Delta\Delta G$ value calculated from the potential of mean force (Fig. 7) suggest that the A-B salt bridge is the most stabilizing of the three possible salt bridge interactions in GPKGDO, presumably due to the favorable geometries of the lysine and aspartate residues in question for adopting this conformation. In contrast, the C-A salt bridge of the same peptide offers

TABLE 1 Relative free energies of salt bridge formation in different collagen-like model peptides

	GPKGEO			GPKGDO		
	$\Delta\Delta G$ (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
A-B	−0.6	−7.9	7.3	−3.6	−17.6	14.0
B-C	−2.4	−4.5	2.1	−2.9	−5.6	2.7
C-A	−1.3	−9.0	7.7	−1.1	−33.6	32.5

a mere 1.1 kcal/mol of stability compared to the metastable state. In this latter interaction pair, the C α atoms of the two residues are brought into close proximity. The length of the lysine side chain as compared to that of aspartate requires unfavorable stereochemical contortions to adopt a salt bridge conformation. This stereochemistry is relaxed at higher distances, and is likely to be responsible for the relative stability of the secondary state rather than intermediate water molecules.

Recently obtained experimental data in which a heterotrimeric model peptide comprised of (POG)₁₀, (PKG)₁₀, and (DOG)₁₀ chains was observed to form in a single triple-helical register in solution (33). In these experiments, (POG)₁₀, (PKG)₁₀, and (DOG)₁₀ peptide chains were combined in a 1:1:1 ratio, heated to 85°C, then cooled to 25°C and incubated overnight at room temperature (34). In principle, this mixture could yield three different homotrimers and six different heterotrimers (not including different possible chain registers). However, a single heterotrimer—one that included one copy of each peptide sequence (deemed the KDO peptide)—was the dominant product (33). Moreover, this heterotrimer was found to exist in only one of six possible chain registers; multidimensional NMR experiments unambiguously identified chain A as the (PKG)₁₀ chain, chain B as the (DOG)₁₀ chain, and chain C as the (POG)₁₀ chain, as well as demonstrated the formation of interchain salt bridges, involving lysine and aspartate residues, between chains A and B (33). Thus, the observed chain register for the heterotrimeric DKO peptide is consistent with our observation that the lysine-aspartate salt bridge formed between chains A and B is the most energetically favorable.

This study was designed to calculate the contribution of an individual salt bridge to triple-helical stability in collagen. A limitation of our work is that we do not consider the energetic gain/cost associated with the formation of several salt bridges in the system. That is, it is difficult to comment on the cost associated with the formation of a C-A pair when an A-B bridge is formed elsewhere in the chain. Nevertheless, these data represent what we believe to be one of the first attempts to rigorously calculate the influence of these interactions on the stability of a fibrillar protein. Future studies will focus on elucidating the interdependence of these interactions and how multiple salt bridges influence fibrillar protein stability.

Our results shed light on the structural basis for the triple-helical stabilization that results in peptides containing these residue pairs. Interchain electrostatic interactions may play an important role in local stabilization of vertebrate triple-helical collagen in regions requiring added stability. Furthermore, it has been noted that a high natural abundance of these residue pairs occurs in viral and bacterial collagen-like proteins (35,36). Prokaryotes lack the prolyl hydroxylase enzyme responsible for hydroxylation of the proline residues in the Y position (37), and may therefore call on other mechanisms to stabilize the triple-helical structure of

such proteins. Viruses infecting hosts that lack this enzyme may require similar alternatives to enhance the stability of collagen-like genes. It may be possible to adopt similar approaches in the construction of synthetic collagens by introducing salt bridges in strategic locations.

Experimental melting temperatures have confirmed the ability of these residue pairs to recover the stability of a pure GPO-containing peptide (10,33). Stability of the triple helix in regions lacking hydroxyproline may rely strongly on the presence of these salt bridges, and their disruption could lead to abnormalities in the structure and metabolism of collagen. A detailed biophysical understanding of the contribution of salt bridges to triple-helical stability could therefore form the basis of detailed sequence comparisons between diseased and healthy individuals in an up-and-coming age of high-throughput genomic sequencing.

SUPPORTING MATERIAL

One figure is available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(10\)00276-6](http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)00276-6).

This work was supported by the National Science Foundation (0745638, 0821391).

REFERENCES

1. Myllyharju, J., and K. I. Kivirikko. 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet.* 20:33–43.
2. Vogel, H. G. 1974. Correlation between tensile strength and collagen content in rat skin. Effect of age and cortisol treatment. *Connect. Tissue Res.* 2:177–182.
3. Byers, P. H., and R. D. Steiner. 1992. Osteogenesis imperfecta. *Annu. Rev. Med.* 43:269–282.
4. Myllyharju, J., and K. I. Kivirikko. 2001. Collagens and collagen-related diseases. *Ann. Med.* 33:7–21.
5. Brodsky, B., and A. V. Persikov. 2005. Molecular structure of the collagen triple helix. *Adv. Protein Chem.* 70:301–339.
6. Persikov, A. V., J. A. M. Ramshaw, ..., B. Brodsky. 2000. Amino acid propensities for the collagen triple-helix. *Biochemistry.* 39: 14960–14967.
7. Bella, J., M. Eaton, ..., H. M. Berman. 1994. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science.* 266:75–81.
8. Shoulders, M. D., and R. T. Raines. 2009. Collagen structure and stability. *Annu. Rev. Biochem.* 78:929–958.
9. Bretscher, L. E., C. L. Jenkins, ..., R. T. Raines. 2001. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* 123:777–778.
10. Persikov, A. V., J. A. M. Ramshaw, ..., B. Brodsky. 2005. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry.* 44:1414–1422.
11. Katz, E. P., and C. W. David. 1990. Energetics of intrachain salt-linkage formation in collagen. *Biopolymers.* 29:791–798.
12. Rainey, J. K., and M. C. Goh. 2004. An interactive triple-helical collagen builder. *Bioinformatics.* 20:2458–2459.
13. Brooks, B. R., C. L. Brooks, 3rd, ..., M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.

14. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.
15. Feig, M., J. Karanicolas, and C. L. Brooks, III. 2004. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph Model.* 22:377–395.
16. Evans, D. J., and B. L. Holian. 1985. The Nose-Hoover thermostat. *J. Chem. Phys.* 83:4069–4074.
17. Ravikumar, K. M., and W. Hwang. 2008. Region-specific role of water in collagen unwinding and assembly. *Proteins.* 72:1320–1332.
18. Torrie, G. M., and J. P. Valleau. 1974. Monte-Carlo free-energy estimates using non-Boltzmann sampling—application to subcritical Lennard-Jones fluid. *Chem. Phys. Lett.* 28:578–581.
19. Roux, B. 1995. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* 91:275–282.
20. Nerenberg, P. S., and C. M. Stultz. 2008. Differential unfolding of $\alpha 1$ and $\alpha 2$ chains in type I collagen and collagenolysis. *J. Mol. Biol.* 382:246–256.
21. Brooks, III, C. L., M. Karplus, and B. M. Pettitt. 1988. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. J. Wiley and Sons, New York, NY.
22. Kumar, S., D. Bouzida, ..., J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem.* 13:1011–1021.
23. Souaille, M., and B. Roux. 2001. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* 135:40–57.
24. Du, Q., D. Beglov, and B. Roux. 2000. Solvation free energy of polar and nonpolar molecules in water: an extended interaction site integral equation theory in three dimensions. *J. Phys. Chem. B.* 104:796–805.
25. Hendsch, Z. S., and B. Tidor. 1993. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* 3:211–226.
26. Makhatadze, G. I., V. V. Loladze, ..., S. T. Thomas. 2003. Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.* 327:1135–1148.
27. Sindelar, C. V., Z. S. Hendsch, and B. Tidor. 1998. Effects of salt bridges on protein structure and design. *Protein Sci.* 7:1898–1914.
28. Horovitz, A., L. Serrano, ..., A. R. Fersht. 1990. Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* 216:1031–1044.
29. Kumar, S., and R. Nussinov. 2001. How do thermophilic proteins deal with heat? *Cell. Mol. Life Sci.* 58:1216–1233.
30. Anderson, D. E., W. J. Becktel, and F. W. Dahlquist. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry.* 29:2403–2408.
31. Strop, P., and S. L. Mayo. 2000. Contribution of surface salt bridges to protein stability. *Biochemistry.* 39:1251–1255.
32. Tissot, A. C., S. Vuilleumier, and A. R. Fersht. 1996. Importance of two buried salt bridges in the stability and folding pathway of barnase. *Biochemistry.* 35:6786–6794.
33. Fallas, J. A., V. Gauba, and J. D. Hartgerink. 2009. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *J. Biol. Chem.* 284:26851–26859.
34. Gauba, V., and J. D. Hartgerink. 2007. Surprisingly high stability of collagen ABC heterotrimer: evaluation of side chain charge pairs. *J. Am. Chem. Soc.* 129:15034–15041.
35. Rasmussen, M., M. Jacobsson, and L. Björck. 2003. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* 278:32313–32316.
36. Xu, Y., D. R. Keene, ..., S. Lukomski. 2002. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J. Biol. Chem.* 277:27312–27318.
37. Mohs, A., T. Silva, ..., B. Brodsky. 2007. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* 282:29757–29765.